Syllabus for MAT 128: Foundations of Data Science

Course Description MAT 128: *4 hours, 3 credit*s. Statistical and computational tools for analyzing data. Acquiring data from multiple sources, techniques for efficiently traversing, storing, and manipulating data. Emphasis on statistical analysis and visualization of real data.

Course Goal: Learn the fundamentals of data science by analyzing real world datasets with Python programming language.

Textbooks:

- *Computational and Inferential Thinking: The Foundations of Data Science* by Ani Adhikari and John DeNero. **https://www.inferentialthinking.com/chapters/intro.html**

Supplemental:

- *Online Statistics: An Interactive Multimedia Course of Study*. Project leader: David M. Lane. http://onlinestatbook.com/2/index.html
- *How to Think Like a Computer Scientist (Python)* by Jeffrey Elkner, Allen B. Downey, and Chris Meyers. http://interactivepython.org/runestone/static/thinkcspy/index.html

Grading: The grading for the course will be based on:

Assignments	30%
Classwork (in-class quizzes and group work)	35%
Final exam	35%

You must take and pass the final exam to pass the course.

Assignments: All assignments should be submitted on Blackboard. **No late assignments are accepted.** Solutions to each assignment will be posted immediately after the due date. You will then have an additional week to correct your submitted assignment, and the corrected assignment will be graded. You can miss up to 5 assignments without affecting your grade (if you turn in all assignments, we will drop the lowest 5 scores).

Classwork: At every class meeting, there will be an in-class quiz or group work based on the lecture notes, reading, submitted assignments, and laboratory exercises. All in-class quizzes and group work will be weighted equally in the classwork grade. If your final exam grade is higher than any of the quiz or group work grades (ex. if you missed a class and have a grade of 0 for that classwork), then your final exam grade will replace that classwork grade.

Final exam: The final exam is required and will be on Thursday December 19 from 1:30-3:30pm. You must pass the final exam to pass the course.

Jupyter Hub: Labs and assignments will be done in a Jupyter Notebook, which can be accessed through Jupyter Hub (under Lehman Apps on **Lehman One Access**). To access Jupyter Hub off

campus (ex. at home), you will have to connect to the Lehman Virtual Private Network (VPN) first (see instructions on course webpage).

Use of Technology & Blackboard: Homework solutions and grades will be posted on Blackboard. The Blackboard system is provided by CUNY to all enrolled students. If you have not accessed Blackboard or are having difficulties, contact Blackboard Support in the Information Technology Division. You can also visit the Help Desk in the Computer Center (first floor, Carman Hall) in person. They can reset passwords and help with simple Blackboard issues.

Honor Code: You are encouraged to work together on solving the homework problems. However, you should write up the solutions on your own. You are responsible for knowing and following Lehman's academic integrity code (available from the Undergraduate Bulletin, Graduate Bulletin, or the Office of Academic Standards and Evaluations). All incidents of cheating and plagiarism will receive a 0 on the test or assignment, and will be reported to the Office of Student Affairs.

Accommodating Disabilities: Lehman College is committed to providing access to all programs and curricula to all students. Students with disabilities who may need classroom accommodations are encouraged to register with the Office of Student Disability Services. For more info, please contact the Office of Student Disability Services, Shuster Hall, Room 238, phone number, 718-960-8441.

Learning Objectives:
At the end of the course, students will be able to:

1) Using real-world datasets, generate and interpret bar charts, line graphs, histograms, and boxplots, as well as compute measures of central tendency and spread.

2) Differentiate between a sample and a population, sample from a distribution, and understand and draw inferences from sampling distributions.

3) Estimate probabilities from real-world datasets.

4) Conduct one and two sample hypothesis tests using simulation.

5) Compute and interpret a regression model for a real-world dataset.

6) Understand and use a nearest neighbour classifier.

Class Schedule:
See course website.